Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2006, Article ID 25861, Pages 1–13
DOI 10.1155/WCN/2006/25861

Multiservice Vertical Handoff Decision Algorithms

Fang Zhu and Janise McNair

Wireless & Mobile Systems Laboratory, Department of Electrical and Computer Engineering, University of Florida, P.O. Box 116130, Gainesville, FL 32611, USA

Received 8 October 2005; Revised 22 March 2006; Accepted 26 May 2006

Future wireless networks must be able to coordinate services within a diverse-network environment. One of the challenging problems for coordination is vertical handoff, which is the decision for a mobile node to handoff between different types of networks. While traditional handoff is based on received signal strength comparisons, vertical handoff must evaluate additional factors, such as monetary cost, offered services, network conditions, and user preferences. In this paper, several optimizations are proposed for the execution of vertical handoff decision algorithms, with the goal of maximizing the quality of service experienced by each user. First, the concept of policy-based handoffs is discussed. Then, a multiservice vertical handoff decision algorithm (MUSE-VDA) and cost function are introduced to judge target networks based on a variety of user- and network-valued metrics. Finally, a performance analysis demonstrates that significant gains in the ability to satisfy user requests for multiple simultaneous services and a more efficient use of resources can be achieved from the MUSE-VDA optimizations.

1. INTRODUCTION

Future wireless networks must be able to coordinate services within a diverse network environment. For example, a widely deployed third generation (3G) cellular and data service, such as the general packet radio service (GPRS), may be supplemented by the local deployment of high bandwidth wireless local area networks (WLANs), such as IEEE 802.11 and the European high performance radio LAN (HiperLAN). Furthermore, as shown in Figure 1, existing networks, such as satellite, cellular, and WLAN, will need to integrate with emerging networks and technologies, such as wireless mesh networks and Wi-Max to allow a user to transparently and seamlessly roam between systems.

Seamless roaming involves handoff, which is the process of maintaining a mobile users active connections as it moves within a wireless network [1]. Vertical handoff, or intersystem handoff, involves handoff between different types of networks [2, 3]. Traditionally, handoff decisions have been based on an evaluation of the received signal strength (RSS) between the base station and the mobile node. However, traditional RSS comparisons are not sufficient to make a vertical handoff decision, as they do not take into account the various attachment options for the mobile user. More recently, bandwidth and the type of network have been considered as factors. For example, the third generation partnership project (3GPP) is currently developing standards for the issue of when, where, and how to initiate a vertical handoff between the 3G cellular network and WLAN networks. Future wireless integration must include still other relevant factors, such as monetary cost, network conditions, mobile node conditions, and user preferences, as well as the capabilities of the various networks in the vicinity of the user. Thus, a complex, adaptive, and intelligent approach is needed to implement vertical handoff protocols to produce a satisfactory result for both the user and the network.

1.1. Related work

Related work on vertical handoff has been presented in recent research literature. Several papers have addressed designing an architecture for hybrid networks, such as the application-layer session initiation protocol (SIP) [4], the hierarchical mobility management architecture proposed in [5], and the *P*-handoff protocol [6], which complemented classical vertical handoff by redirecting traffic to the best ad hoc link, such as Bluetooth and 802.11b, on a peer-by-peer basis. However, these papers focused on architecture design and did not address the handoff decision point or the vertical handoff performance issues. Another work considered optimizations after the vertical handoff decision has been made, measuring performance with respect to handoff latency [7],

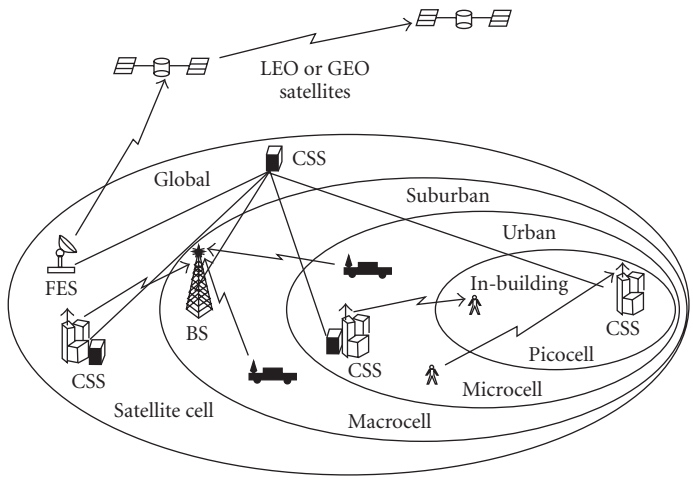


Figure 1: Diverse third- and fourth-generation (3G and 4G) wireless networks.

TCP timeout and throughput [8, 9], and packet loss [10]. However, the vertical handoff decision did not consider multiple networks supporting multiple services for each user.

The related papers that explored vertical handoff decision mainly focus on traditional issues, such as RSS and data rate. In [11], a fast-Fourier-transform- (FFT-) based signal decay detection scheme was used to reduce the ping-pong handoff effect, and an adaptive threshold configuration approach was proposed to prolong the time a user stays in WLAN. In [12, 13], a vertical handoff algorithm was proposed that took into account RSS, data rate, and packet loss due to handoff delay for a single service per user. A vertical handoff system based on computed background noise and signal strength was proposed in [14]. In [15], the WISE handoff decision algorithm was proposed to maximize energy-efficiency without sacrifice of overall network degradation. In [16], a QoS-based handoff method between UMTS and WLAN was proposed, but the definition of QoS was not defined in the paper. Finally, several papers have focused on mobility level and user position in the network. In [17], mobility level was proposed as a proper metric for multi-tier handoffs. In [2], multi-network architectural issues were explored, and an advanced neural-network-based vertical handoff algorithm was developed to satisfy user bandwidth requirements. In [18], a vertical handoff algorithm based on pattern recognition was presented. Although the above-mentioned research addresses handoff decision, most research address 3G/WLAN issues, and do not provide a way to incorporate a general, user-defined idea of quality of service, on which to base vertical handoff decisions.

Several papers have created utility functions to better evaluate the choice for vertical handoff. In [19], the vertical handoff decision function was a measurement of network quality. However, no performance analysis was provided. In [20], an active application-oriented handoff decision algorithm was proposed for multi-interface mobile terminals to reduce the power consumption caused by unnecessary handoffs and other unnecessary interface activation, and in [21], a policy-enabled handoff decision algorithm was proposed along with a cost function that considers several handoff metrics. However, multi-service handoff was not fully discussed. However, the multiple active services case was not considered. The work in [22] adaptively adjusted the handoff stability period based on a utility function to avoid unnecessary handoffs and reduce decision time. Finally, the authors have presented a tutorial on vertical handoffs in [3], and in [23], introduce a cost function-based vertical handoff decision algorithm for multiservices handoff. Preliminary results demonstrated significant gain in throughput. This paper extends the work to examine the system performance with respect to blocking probability and user satisfactions, that is, the ability of the network to satisfy all of the users simultaneous requests.

In this paper, several optimizations are proposed to enhance the handoff decision process and to make the following contributions: (1) the development of a handoff cost function that addresses an environment where users conduct multiple active sessions among a variety of wireless network choices, (2) the design of a multiservice vertical handoff decision algorithm (MUSE-VDA), which incorporates a network elimination process to potentially reduce delay and processing in the handoff calculation, and (3) a constraint optimization analysis for the proposed handoff cost function for different types of user services spread among multiple networks. In Section 2, the policy-based handoff approach is described. Section 3 introduces the MUSE-VDA cost function and algorithm to decide target networks based on a variety of user- and network-valued metrics. Finally, in Sections 4 and 5, the performance analysis and numerical results demonstrate the load-balancing advantages of the proposed technique, as well as the significant gains in satisfied user requests and a more efficient use of resources. Section 6 concludes the paper.

2. POLICY-BASED VERTICAL HANDOFFS

Vertical handoff performed on a policy-based networking architecture requires the coordination of a wide variety of

network devices within a single administrative domain to implement a set of quality-of-service- (QoS-) based services [24]. Figure 2 shows two possible conceptual architectures of policy-based solutions that have been proposed by the IETF. The two main architectural elements for policy control are the policy enforcement point (PEP) and the policy decision point (PDP). These two elements may be located in the same network node (as shown in Figure 2(a)) or in different nodes (as shown in Figure 2(b)). The latter is especially convenient to apply local policies.

PEP is a component that runs on a policy-aware node, such as an access point, and is the point at which the policies are enforced. Policy decisions are made primarily at the PDP, based on the policies extracted from a network policy database. The PDP as specified by the IETF may make use of additional mechanisms and protocols to achieve additional functionality such as user authentication, accounting, and policy information storage.

In the case of vertical handoff, the policy database holds information regarding the metrics to be considered for a vertical handoff, where handoff metrics are the measured qualities that give an indication of whether or not a handoff is needed. As stated previously, in traditional handoffs, only RSS and channel availability are considered. In the envisioned integrated wireless system, the following new metrics are suggested [3].

(i) *Service type*. Different types of services require various combinations of reliability, latency, and data rate.

(ii) *Monetary cost*. A major consideration to users, as different networks may employ different billing strategies that may affect the user's choice to handoff.

(iii) *Network conditions*. Network-related parameters such as traffic, available bandwidth, network latency, and congestion (packet loss) may need to be considered for effective network usage. Use of network information in the choice to handoff can also be useful for load balancing across different networks, possibly relieving congestion in certain systems.

(iv) *System performance*. To guarantee the system performance, a variety of parameters can be employed in the handoff decision, such as the channel propagation characteristics, path loss, interchannel interference, signal-to-noise ratio (SNR), and the bit error rate (BER). In addition, battery power may be another crucial factor for certain users. For example, when the battery level is low, the user may choose to switch to a network with lower power requirements, such as an ad hoc Bluetooth network.

(v) *Mobile terminal conditions*. MT condition includes dynamic factors such as velocity, moving pattern, moving histories, and location information.

(vi) *User preferences*. User preference can be added to cater to special requests for users that favor one type of system over another.

The use of new vertical handoff metrics and the policy-based networking architecture increases the complexity of the handoff process and makes the handoff decision more and more ambiguous. However, the use of an optimized cost function can simplify the handoff process and speed up the handoff decision. Then, intelligent techniques can be developed to evaluate the effectiveness of new decision algorithms, balanced against user satisfaction and network efficiency.

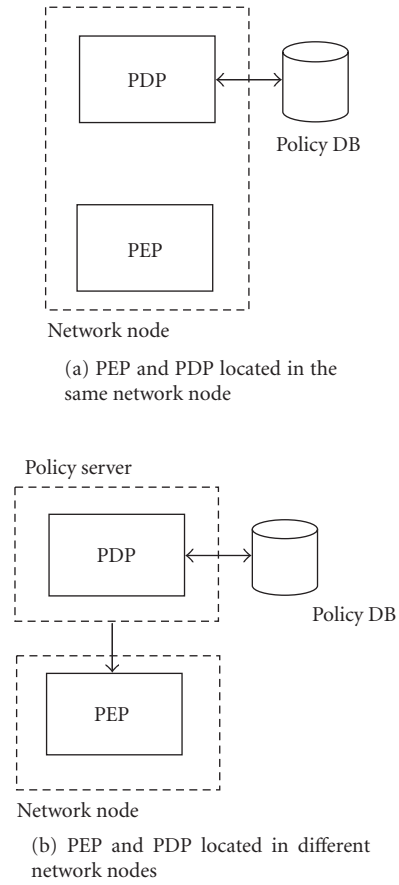
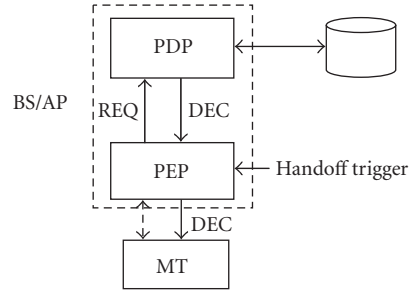


Figure 2: Two possible policy-based network architectures.

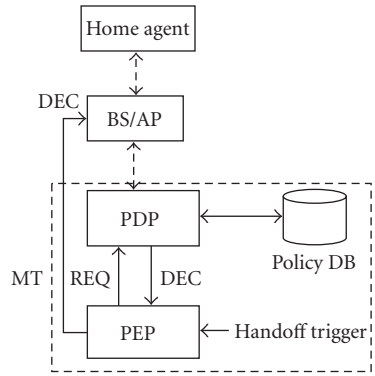
2.1. Proposed vertical handoff interworking scenarios

To demonstrate the operation of the policy-based architectures, the following two scenarios are described: (1) network-controlled handoff (NCHO)/mobile-assisted handoff (MAHO), where the network generates a new connection and finds new resources for the handoff, performing any additional routing operations, and (2) mobile-controlled handoff (MCHO), where the mobile terminal must take its own measurements and make the evaluations for the handoff decision.

NCHO/MAHO is shown in Figure 3(a). The handoff decision procedure begins with the PEP. Upon receiving a handoff trigger, the PEP formulates a request for a policy decision and sends it to the PDP. The request for policy control from the PEP to the PDP may contain one or more policy elements extracted from the mobile terminals that are necessary for handoff decision. The PDP then extracts other necessary information, for example, the users subscriber profile



(a) NCHO or MAHO handoff decision procedure



(b) MCHO handoff decision procedure

Figure 3: Two scenarios for policy-based architectures.

and network conditions, from the database located in local or home network, makes the handoff decision, and returns the decision message to the PEP. The handoff decision is made using utility-function-based algorithms as proposed in [23]. The PEP then informs the mobile terminal about the handoff decision and enforces the policy decision by handing off to the target network. In NCHO/MAHO, we propose that the PDP point is represented by the base station (BS) or access point (AP).

In MCHO, the mobile terminal finds new resources and the network approves the handoff decision. Thus, we propose that the PDP is located at the mobile terminal. As shown in Figure 3(b), when the mobile terminal detects a severe QoS degradation, its PEP module triggers the handoff decision process by sending a handoff decision request message to the PDP. While some information is already available at local database, the PDP may also need other necessary information, such as network conditions, from the network devices. Other information may not be immediately available at the BS or AP, and may need to be extracted from the network. Upon receiving all handoff metrics, the PDP makes the handoff decision and returns the decision to the PEP. The PEP then informs the network the handoff decision by forwarding the DEC message, along with enforced authentication information. A handoff will take place once the network approves. It may be a limiting factor to achieve the necessary processing for a vertical handoff controlled by the mobile terminal. However, if simple metrics are set, a combination of the two techniques, that is, mobile-assisted handoff (MAHO), may be a viable option.

3. MULTISERVICE VERTICAL HANDOFF DECISION ALGORITHM COST FUNCTION

The MUSE-VDA vertical handoff cost function measures the benefit obtained by handing off to a particular network. It is evaluated for each network n that covers the service area of a user. The network choice that results in the lowest calculated value of the cost function is the network that provides the most benefit, where the benefit is defined by the given handoff policy.

The cost function evaluated for network n includes the cost of receiving each of the user's requested services from network n and is calculated:

$$C^n = \sum_s C_s^n, \tag{1}$$

where s is the index representing the user-requested services, and C_s^n is the per-service cost function for network n. C_s^n represents the QoS experienced by choosing to receive service s from network n and is calculated as

$$C_s^n = \sum_j W_{s,j}^n Q_{s,j}^n, \tag{2}$$

where $Q_{s,j}^n$ is the normalized QoS provided by network n for parameter j and service s. $W_{s,j}^n$ is the weight which indicates the impact of the QoS parameter on the user or the network. C_s^n includes both a normalized value for the QoS parameter and a weight for the impact of the parameter on either the user or the network. For an example from the users perspective, suppose that a mobile terminal requests a service with a specified minimum delay and minimum power consumption requirement. If the mobile terminal has a low battery life, the power consumption takes on greater importance than meeting the delay constraints. For an example of a network-based QoS request and the corresponding impact, the availability of the services requested by the user in the target network impacts the network congestion in the target network. Using the impact factor, the network may direct users toward a less desirable, but less congested network.

The handoff decision problem thus equals the following constraint optimization problem:

$$\min C^n = \sum_s C_s^n \sum_j W_{s,j}^n Q_{s,j}^n \quad \text{s.t. } E_{s,j}^n \neq 0, \quad \forall s, i, \tag{3}$$

where $E_{s,j}^n$ is the network elimination factor, indicating whether the constraint i for service s can be met by network n. It is equal to one if constraint i can be satisfied, and is equal to zero if constraint i cannot be satisfied. It is introduced to reflect the inability of a network to guarantee the requested QoS constraints for a particular service s, and can be implemented as a checklist at PDP. For example, an available network may not be able to guarantee the minimum

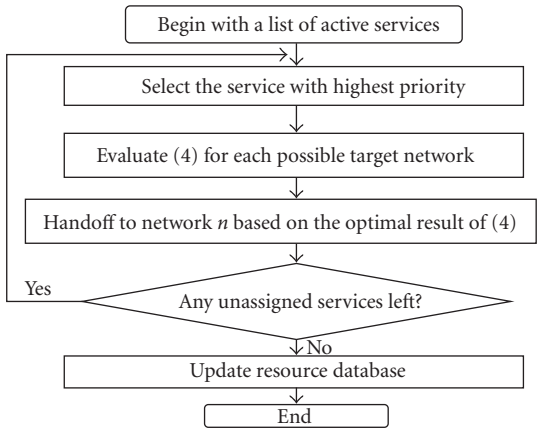


Figure 4: Scenario 2: prioritized session handoff.

requested delay for a real-time service, and should be immediately removed from consideration as a handoff target for the requested service.

The application of the vertical handoff cost function is flexible to allow for different vertical handoff policies. To demonstrate the performance of the new cost function, two different policy scenarios are explored.

3.1. Collective session handoff

It is assumed that a single user may conduct multiple communication sessions. In the first vertical handoff policy, the vertical handoff decision is optimized for all sessions collectively, that is, all of the users active sessions are handed off to the same target network at the same time. The cost function, C^n, is determined for all sessions going to a single network. The optimal target network for handoff is determined by solving (3).

3.2. Prioritized session handoff

The second vertical handoff policy prioritizes each service and then optimizes the vertical handoff decision individually for each session, that is, each of the users active sessions may be independently handed off to a different target network. In this scenario, the mobile terminal maintains a list of its current active sessions, arranged in priority order. Then, the cost function C_s^n is evaluated for the highest priority service. The optimal target network is chosen by minimizing the per-service cost:

$$\min C_s^n = \sum_s W_{s,j}^n Q_{s,j}^n \quad \text{s.t. } E_{s,j}^n \neq 0, \quad \forall i. \tag{4}$$

Then, the next highest priority service is selected, the corresponding cost function is evaluated, and the target network is determined. The process continues to the last active session. If the constraints for one session cannot be met, then the user loses the individual session only. The process for the second scenario is outlined in Figure 4.

3.3. Cost function example

As an example, consider a reporter in the field using wireless networks to send audio, video reports, and photographic images to a home base, but whose equipment is running low on battery power. There are three available networks, UMTS, WLAN, and satellite. The cost function calculation from (3) is formed as follows:

(i) n represents the three network choices, UMTS, WLAN, or a satellite network.
(ii) s represents the services needed, in this case, audio, video, and images.
(iii) j represents the constraint parameters: bandwidth, battery power consumption, and delay.
(iv) For collective handoff, a calculation of (3) is made for each network.

(1) For example, for the UMTS network,

$$\begin{aligned} C^{\text{UMTS}} = {} & \left(W_{\text{video, bandwidth}}^{\text{UMTS}} Q_{\text{video, bandwidth}}^{\text{UMTS}} \right. \\ & + W_{\text{video, battery power}}^{\text{UMTS}} Q_{\text{video, battery power}}^{\text{UMTS}} \\ & \left. + W_{\text{video, delay}}^{\text{UMTS}} Q_{\text{video, delay}}^{\text{UMTS}}\right) \\ & + \left(W_{\text{audio, bandwidth}}^{\text{UMTS}} Q_{\text{audio, bandwidth}}^{\text{UMTS}} \right. \\ & + W_{\text{audio, battery power}}^{\text{UMTS}} Q_{\text{audio, battery power}}^{\text{UMTS}} \\ & \left. + W_{\text{audio, delay}}^{\text{UMTS}} Q_{\text{audio, delay}}^{\text{UMTS}}\right) \\ & + \left(W_{\text{image, bandwidth}}^{\text{UMTS}} Q_{\text{image, bandwidth}}^{\text{UMTS}} \right. \\ & + W_{\text{image, battery power}}^{\text{UMTS}} Q_{\text{image, battery power}}^{\text{UMTS}} \\ & \left. + W_{\text{image, delay}}^{\text{UMTS}} Q_{\text{image, delay}}^{\text{UMTS}}\right). \end{aligned} \tag{5}$$

(2) Then, C^{WLAN} and $C^{\text{Satellite}}$ are calculated similarly.
(3) The lowest of the three costs C^{UMTS}, C^{WLAN}, and $C^{\text{Satellite}}$ reveals the target network. If satellite cost is the lowest, then all sessions, video, audio, and images, are sent via the satellite network.

(v) For prioritized session handoff, a calculation of (4) is made for the highest priority session.

(1) For example, if the video feed has the highest priority, then $C_{\text{video}}^{\text{UMTS}}$ is calculated first:

$$\begin{aligned} C_{\text{video}}^{\text{UMTS}} = {} & \left(W_{\text{video, bandwidth}}^{\text{UMTS}} Q_{\text{video, bandwidth}}^{\text{UMTS}} \right. \\ & + W_{\text{video, battery power}}^{\text{UMTS}} Q_{\text{video, battery power}}^{\text{UMTS}} \\ & \left. + W_{\text{video, delay}}^{\text{UMTS}} Q_{\text{video, delay}}^{\text{UMTS}}\right). \end{aligned} \tag{6}$$

(2) Then $C_{\text{video}}^{\text{WLAN}}$ and $C_{\text{video}}^{\text{Satellite}}$ are calculated similarly.
(3) The lowest of the three costs $C_{\text{video}}^{\text{UMTS}}$, $C_{\text{video}}^{\text{WLAN}}$, and $C_{\text{video}}^{\text{Satellite}}$ reveals the target network *for video service only*.

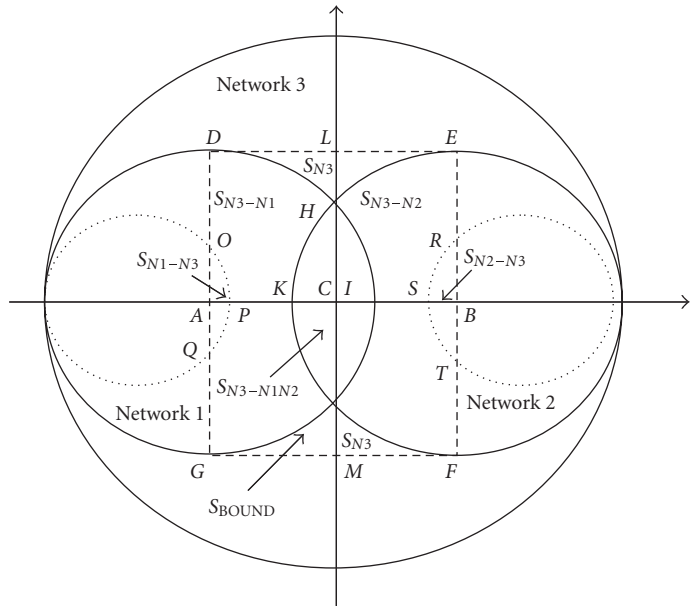


Figure 5: 3G/WLAN overlay network scenario.

(4) The calculation is repeated for the next highest priority service, say the audio feed. Thus, in the prioritized session handoff it may be the case that the video is sent via satellite for the bandwidth, but the audio is sent via UMTS.

In the next section, the performance of the proposed MUSE-VDA algorithm and cost function is analyzed. First, a sample overlay network scenario is provided, along with a description of the mobility model, followed by calculations of the blocking probability and the average percentage of user requests that are satisfied by the network.

4. MUSE-VDA PERFORMANCE ANALYSIS

For effective comparison with other techniques, the performance analysis considers the case of 3G/WLAN handoff scenario, where received signal strength (RSS), channel availability, and bandwidth are the specified constraints. However, note that any other network combination or any other combination of the vertical handoff metrics listed in Section 2 can just as easily be substituted in the evaluation.

The top view of a typical 3G network overlay environment is shown in Figure 5, where three networks of different maximum data rates coexist in the same wireless service area. Network 1 (centered at A) and Network 2 (centered at B) each represent a WLAN, while Network 3 (centered at C) represents a GPRS network. The shaded circles on the left and right represent the area where RSS from Network 1 or Network 2 is stronger than that from Network 3. To highlight the effects of the vertical handoff procedure among the three networks, only the users within the overlapping areas are considered, represented by the dashed square in Figure 5.

4.1. Mobility model

User mobility trajectories are characterized by the widely used random waypoint (RWP) model [25]. Adjustments have been included to account for the shortcomings of the waypoint model described in [12]. Each user chooses uniformly at random a destination point (or waypoint) in the dashed rectangle in Figure 5. A user moves to this destination with a velocity v, which is chosen uniformly in the interval (v min, v max 0). (The v min and v max are chosen to be 0.3 m/s and 12.5 m/s, resp.) When the user reaches the waypoint, it remains static for a predefined pause time, and then moves again according to the same rule. Note that user trajectories characterized by the improved RWP model can be assumed to be uniformly distributed at any given time.

A user with active sessions that enters the overlay of all three networks must decide when and where to execute a vertical handoff request. If the request is accepted, the appropriate amount of bandwidth is assigned by the serving network. If the request is denied at one network, the request can be reassigned to another network, if resources are available at the second network. If the second (or third) network is not available, the request is blocked from the system. Next, we formulate the calculation of the blocking probabilities.

4.2. Blocking probability

Each of the three networks in Figure 5 is modeled as an $M/M/1/N_n$ queue system [26], where N_n is the number of available channels in Network n. N_n is calculated:

$$N_n = \frac{B_n}{D}, \tag{7}$$

where B_n is the total bandwidth of Network n, and D is the average data rate of each user. The traffic load within the overlay cells is $\rho = \lambda/\mu$, where λ is the arrival rate of service requests, μ is the departure rate, and arrivals and departures are modeled as Poisson distributions. Handoff calls are given a higher priority than new calls, and for simplicity, a bufferless handoff algorithm is used.

For the blocking probability of Network n, P_{bn}, we use the blocking probability of an $M/M/1/N_n$ queue when there are N_n users in system [26]:

$$P_{bn} = \frac{\rho_n^{N_n}(1-\rho_n)}{1-\rho_n^{N_n+1}}, \tag{8}$$

where ρ_n is the effective load experienced by Network n:

$$\rho_n = r_n \rho \tag{9}$$

and r_n is the percentage of total requests that will go to Network n, based on the vertical handoff decision metrics. To determine r_n, both original handoff requests and the handoff requests that arrive are included, to account for the times that the user has been rejected by another network. Since it is assumed that the users are uniformly distributed, the service request load can be calculated according to the proportion of the coverage area within the boundary region. The coverage areas are labeled in Figure 5, and the corresponding coverage, the execution of the RSS and MUSE-VDA algorithms are described in Table 1.

For the RSS-based handoff algorithm, the values of r_n for $n = 1, 2, 3$ are calculated as follows:

$$\begin{aligned} r_3 &= \frac{S_{\text{BOUND}} - S_{N1-N3} - S_{N2-N3}}{S_{\text{BOUND}}}, \\ r_1 &= \frac{S_{N1-N3} + S_{N3-N1} P_{b3}}{S_{\text{BOUND}}}, \\ r_2 &= \frac{S_{N2-N3} + S_{N3-N2} P_{b3}}{S_{\text{BOUND}}}, \end{aligned} \tag{10}$$

where P_{b3} is defined in (8), S_i is the geometric area of region i described in Table 1, and S_{BOUND} is the geometric area of the boundary region.

For the MUSE-VDA handoff algorithm, the values of r_n for $n = 1, 2, 3$ are calculated:

$$\begin{aligned} r_1 &= \frac{S_{N1-N3} + S_{N3-N1} + S_{N3-N1N2}}{S_{\text{BOUND}}}, \\ r_1 &= \frac{S_{N2-N3} + S_{N3-N2} + S_{N3-N1N2} P_{b1}}{S_{\text{BOUND}}}, \\ r_2 &= \frac{S_{N3} + (S_{N1-N3} + S_{N3-N1}) P_{b1} + (S_{N2-N3} + S_{N3-N2} + S_{N3-N1N2}) P_{b2}}{S_{\text{BOUND}}}. \end{aligned} \tag{11}$$

Finally, we develop a calculation for a measure of the service obtained by each user, as compared to the services requested by each user. This is defined here as *average percentage of users' satisfied requests* (APUSR).

4.3. Average percentage of satisfied user requests

Each user comes to the network overlay area with a certain set of requests, including various services and data rates. As mentioned previously, the ability of the network to satisfy user requests depends on whether the sessions are treated as a collective or as prioritized, individual sessions. In the collective MUSE-VDA and the RSS technique, all requests from one user are considered collectively. Thus, if a target network cannot satisfy all of the requests as a collective, then the user is blocked from the system. In the prioritized MUSE-VDA technique, each session is treated individually, and thus one user may have a subset of their requests satisfied, while other portions are blocked. The APUSR tracks the percentage of incoming requests that actually receive service at one of the available networks.

The APUSR is calculated for the overlay network as follows:

$$E[A_R] = \sum_i A_{R_i} P(R_i), \tag{12}$$

where A_{R_i} is the APUSR for Region i, and where the regions are described in Table 1. A_{R_i} is calculated:

$$A_{R_i} = \sum_j t_{ij} P(N_{ij}), \tag{13}$$

where t_{ij} is the maximum APUSR that can be received from Network N_{ij} in Region i, and $P(N_{ij})$ is the probability that Network N_{ij} is available and chosen by a user. Finally, $P(R_i)$ is the probability that a user is located in Region i:

$$P(R_i) = \frac{S_i}{S_{\text{BOUND}}}. \tag{14}$$

In the next section, we implement the performance analysis and obtain results for several service request scenarios.

5. NUMERICAL RESULTS

The user mobility, user requests, network acceptances and denials for the 3G/WLAN overlay system in Figure 5 were modeled and simulated using MATLAB, based on the system parameters shown in Table 2. Each user can request a data rate up to a maximum of 500 kbps. To gauge the response of the protocol to different traffic types, this data rate includes a combination of constant bit rate (CBR) services and available bit rate (ABR) services, where the CBR request per user is limited to a maximum of 50 kbps and the ABR request per user is limited to a maximum of 450 kbps. Note that Network 1 or Network 2 can fully satisfy the maximum possible data rate request of 500 kbps. However, Network 3 can only satisfy 30% of the maximum possible 500 kbps request. We note that the data rates for the networks listed in Table 2 can be considered as low estimates. However, the objective is to gauge the ability of a combination of networks to satisfy as many user requests as possible. Thus, as data rates per network increase, the size of the data rate request may also increase, but the resulting trends for the given algorithms would remain the same.

Table 1: RSS and MUSE-VDA algorithm description for 3G WLAN overlay network in Figure 5.

Region number	Region area label	Algorithm descriptions
1	S_{N3}	(DEH,JFG) Network 3 provides the only coverage
2	$S_{N3-N1N2}$	(HIJK) Network 3 has the strongest RSS. *RSS Algorithm*: if the request is denied by Network 3, the user can try either Network 1 or Network 2 with equal probability. *MUSE-VDA*: the network order with respect to decreasing data rate is as follows: Network 1 > Network 2 > Network 3. The outcome of the cost function will be to choose Network 1, then Network 2 if Network 1 is denied, then Network 3, if Network 2 is denied.
3	S_{N3-N1}	(DHKJGP) Network 3 has the strongest RSS. *RSS Algorithm*: Network 3 is chosen first. If the request is denied by Network 3, the user tries Network 1. *MUSE-VDA*: according to the decreasing data rates, the selection made by the cost function is first Network 1, then Network 3 if Network 1 is denied.
4	S_{N3-N2}	(EHIJFS) Network 3 has the strongest RSS. *RSS Algorithm*: Network 3 is chosen first. If the request is denied by Network 3, the user tries Network 2. *MUSE-VDA*: according to the decreasing data rates, the selection made by the cost function is first Network 2, then Network 3 if Network 2 is denied.
5	S_{N1-N3}	(OPQA) Network 1 has the strongest RSS. *RSS Algorithm*: Network 1 is chosen first. If the request is denied by Network 1, the user tries Network 3. *MUSE-VDA*: according to the decreasing data rates, the selection made by the cost function is first Network 1, then Network 3 if Network 1 is denied.
6	S_{N2-N3}	(RSTB) Network 2 has the strongest RSS. *RSS Algorithm*: Network 2 is chosen first. If the request is denied by Network 2, the user tries Network 3. *MUSE-VDA*: according to the decreasing data rates, the selection made by the cost function is first Network 2, then Network 3 if Network 2 is denied.
	S_{bound}	Boundary region

Table 2: System parameters.

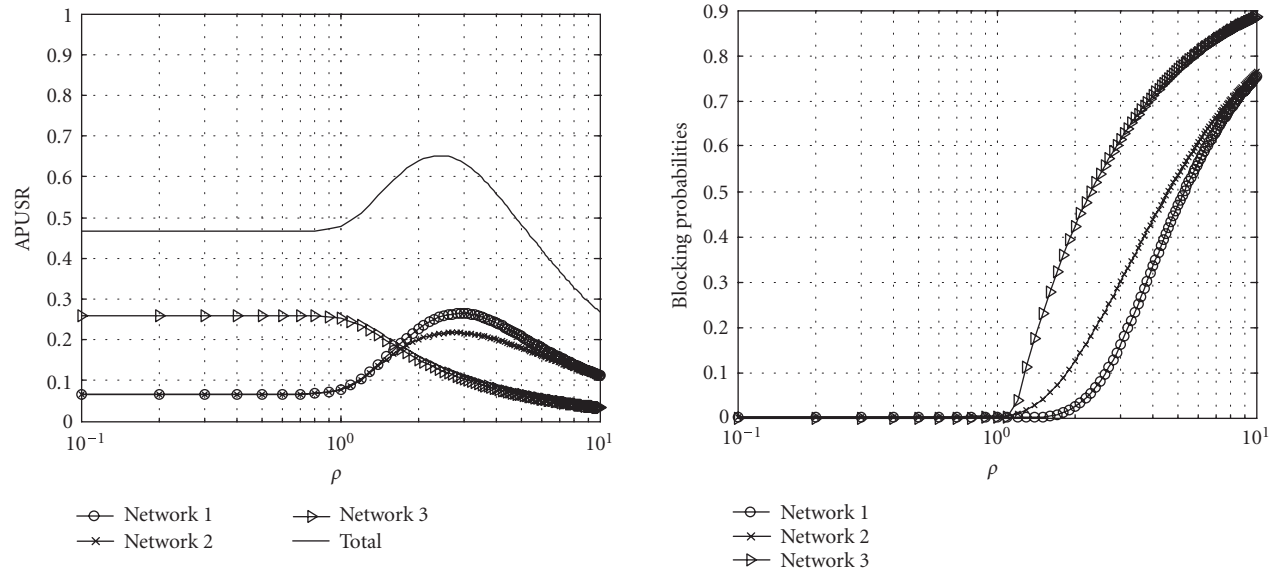
Network (n)	Network type	Resource
1	WLAN	2 Mbps [27]
2	WLAN	1 Mbps [27]
3	GPRS	Up to 8 slots, 21.4 Kbps per slot [27]

As mentioned previously, the random waypoint model is used to simulate user mobility, with the following parameters: $v_{\min} = 0.3$ m/s (1 km/h), $v_{\max} = 12.5$ m/s (45 km/h), and $v_{\text{threshold}} = 5.5$ m/s (20 km/h).

5.1. RSS-based algorithm results

First, the RSS performance is examined to provide a baseline for comparison with the MUSE-VDA results. Figure 6(a) shows the APUSR with the increasing network load for an RSS-based handoff algorithm. Since Network 3 has the strongest transmit power, it is the preferred service provider. Thus, at the low-load range, Network 3 must satisfy a large portion of the total requests. With increasing network load, the resources of Network 3 are used up earlier than the resources of the other two networks. The affect is to separate the APUSR into three regions.

(1) In the first region, $0.1 < \rho < 1$, most of the requests go to GPRS (Network 3), while the WLANs are underused.



(a) Average percentage of user satisfied requests (APUSR)

(b) Blocking probability

Figure 6: Performance of the RSS-based algorithm.

(2) In the second region, $1 < \rho < 2$, GPRS begins to deny users, and the WLANs begin to receive more requests.
(3) In the third region, $2 < \rho$, all three networks are saturated and the QoS degrades for all networks.

Thus, the problem with the RSS approach is that there is no load balancing according to the service requests of the users and the available networks.

Figure 6(b) demonstrates the corresponding blocking probability of each network for the traditional RSS algorithm. An increase in blocking probability of Network 3 earlier than Networks 1 and 2 can be observed. Mobile users thus have a greater chance to select Network 1 and Network 2 as service provider. Since they have a total APUSR that is higher than Network 3 by itself, a "hump" can be observed. The result that Network 3 is chosen more often as the target handoff cell leads to two unsatisfactory effects: (1) unbalanced load assignment and (2) low overall achievable data rate. Only when the resource in Network 3 is highly consumed, Networks 1 and 2 will have a greater chance to be the service provider. Thus a more intelligent handoff algorithm that can balance the usage of overlay networks is needed, and a higher overall APUSR is expected.

5.2. RSS with mobility metric

Next, we compare the RSS-only technique versus a mobility-level technique. Mobility level is a metric that can be combined with RSS based to improve system performance. For example, fast moving users ($v > v_{\text{threshold}}$) are selected to receive service from the largest cell, while medium-to-slow users ($v < v_{\text{threshold}}$) receive service from the small cells. Figure 7 shows the APUSR and blocking probability comparison of the pure RSS based algorithm and the RSS-based algorithm combined with mobility level consideration. The mobility level algorithm demonstrates an improved APUSR performance. However, its achievable APUSR is lower than that of MUSE-VDA (which will be discussed in more detail later in this section), that is, there remains a load-balancing issue for increasing requests.

We now examine the MUSE-VDA performance by considering two handoff scenarios: (1) collective handoff, where all of the user's active sessions are handed off to the same target network at the same time, and (2) prioritized multinetwork handoff, where each service is prioritized and optimal decision is made individually for each session.

5.3. MUSE-VDA

The MUSE-VDA cost functions, (3) and (4), are evaluated for each network based on the following parameters:

(i) Network index n represents the two WLANs and one GPRS network, as shown in Table 2.

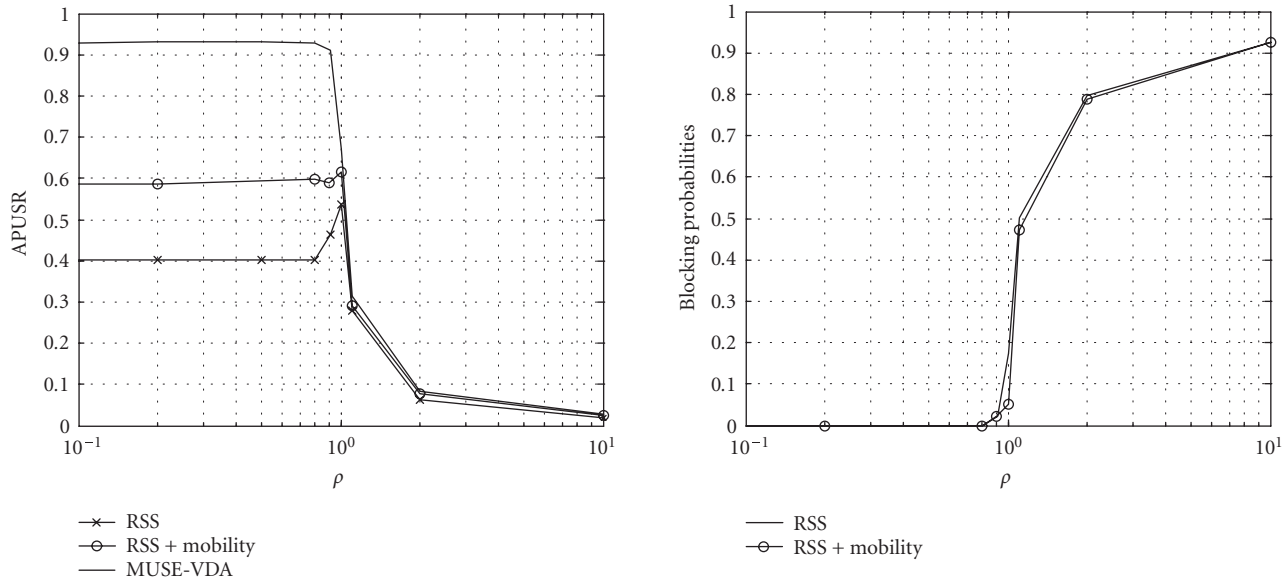
(ii) Two constraints are considered: available bandwidth and RSS (R), where the limiting constraint for bandwidth is $B_s^n - B_{\text{req}} \geq 0$ for some network n and service s, and the limiting RSS contraint is $R^n - R_{\text{th}} \geq 0$.

(iii) The weights in the cost functions are normalized to 1, meaning that each service contraint is treated with equal weight.

(iv) The QoS factor is a normalized bandwidth calculation, where $Q^n_{\text{CBR, bandwidth}} = \ln|1/B^n_{\text{CBR}}|$, and $Q^n_{\text{ABR, bandwidth}} = \ln|1/B^n_{\text{ABR}}|$.

(v) The target network is chosen according to the procedure described in Table 1.

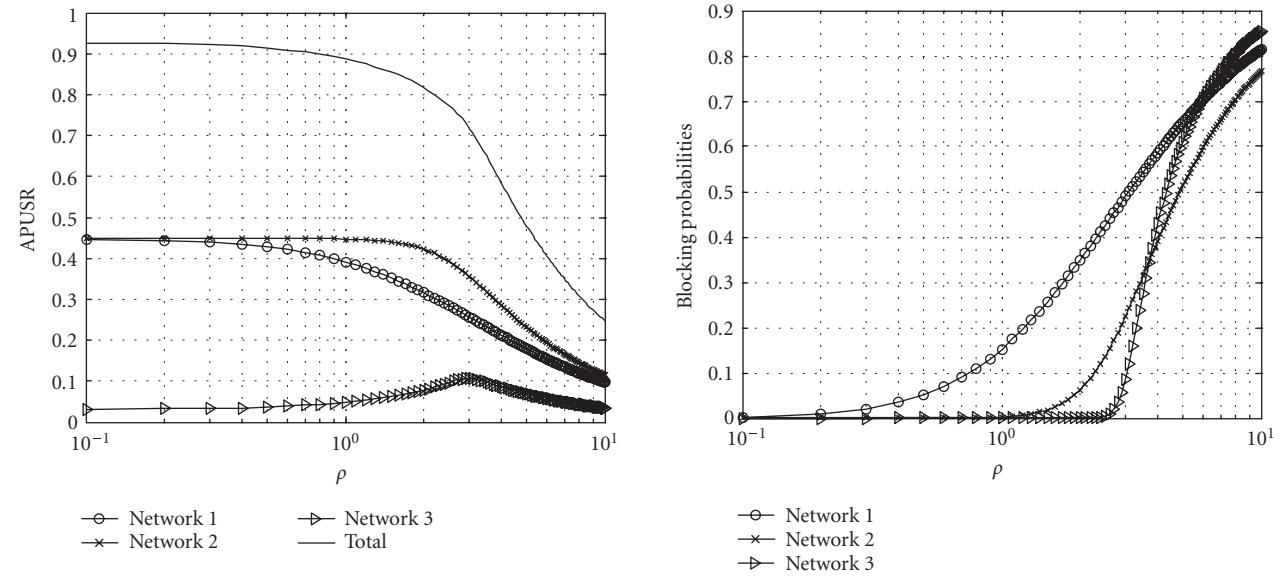
Figure 8(a) shows MUSE-VDA results for the APUSR provided by each of the three networks and overall achievable APUSR implementing the collective handoff algorithm, for comparison with Figure 6, the RSS-only case. Since either Network 1 or Network 2 provides relatively larger data rate than Network 3, they are the default service provider for the



(a) Average percentage of user satisfied requests (APUSR)

(b) Blocking probability

Figure 7: Performance of the RSS-based algorithm with added mobility considerations.



(a) Average percentage of user satisfied requests (APUSR)

(b) Blocking probability

Figure 8: Performance of the MUSE-VDA algorithm.

mobile users, depending on their location. Thus, at the low-load range, Network 1 and Network 2 satisfy the most portion of the total request. With the increasing network load, the resource of Network 1 and Network 2 is consumed earlier than the resources of Network 3. Then mobile users start to select Network 3 more frequently than in low-load range. The portion of requests satisfied by Network 3 thus starts to increase when the portion satisfied by Network 1 and Network 2 decreases. In this case, there are only two regions represented in the figure.

(1) In the first region, $0.1 < \rho < 1$, most of the requests go to the WLANs, which are able to handle the higher data rate requests.

(2) In the second region, $1 < \rho$, WLANs begin to deny users, and the GPRS provides a useful alternative. All three networks are being utilized and the performance degrades gradually.

Thus, in the MUSE-VDA case, the load balancing is improved for all networks.

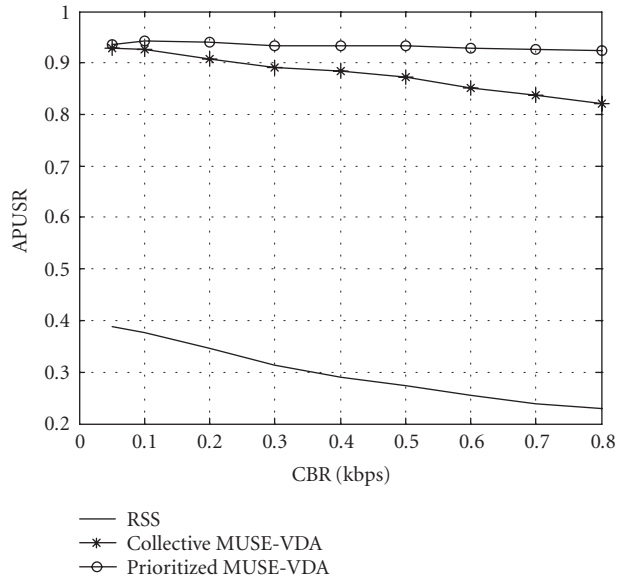


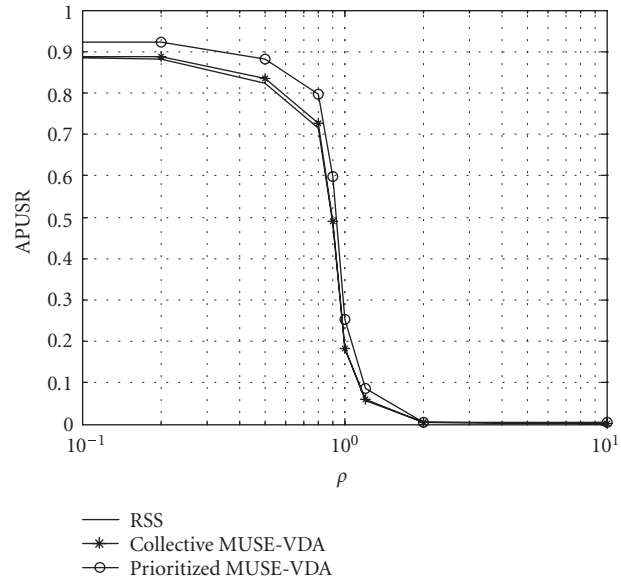
Figure 9: APUSR with varying CBR data rate requests.

Figure 8(b) demonstrates the corresponding blocking probability of each network. An increase in blocking probability of Networks 1 and 2 earlier than Network 3 can be observed, which indicates that WLANs are favorite networks due to their relative larger available bandwidth to each user.

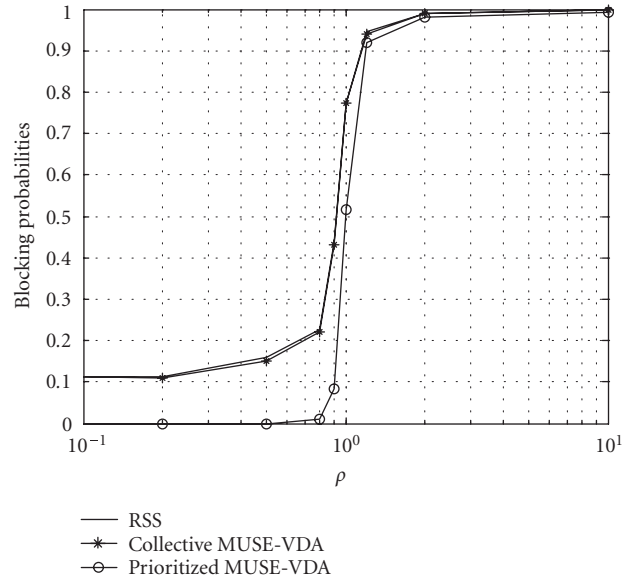
5.4. MUSE-VDA results for more demanding CBR services

In future wireless networks, users may request much higher CBR service rates, as video and audio conferencing and other real-time services become prevalent. Thus, the next set of simulations study the impact of increasing the request of CBR services. Results are now presented for APUSR and blocking probability for three cases: the traditional handoff protocol based on the strongest RSS, the cost function with collective handoff, and the cost function with the prioritized handoff. Figure 9 shows APUSR versus CBR (per user) with zero blocking rates for all three algorithms. Here, each user may request variable CBR and up to 1 Mbps ABR services. The figure demonstrates that APUSR decreases with an increase in CBR request for all three algorithms. However, as each user's request increases, the ability of the large network to support the variable data rate decreases more dramatically. By optimally spreading user's services over several networks using the prioritized MUSE-VDA technique, more bandwidth can be assigned to ABR services, which results in a higher overall APUSR per user.

Figure 10 shows APUSR and blocking probability versus user requests for the three handoff algorithms for the more demanding CBR requests. In this case, Network 3 is eliminated in RSS-based and collective MUSE-VDA handoff algorithms, due to its limited data rate per user (less than 170 kbps). Thus, users will only be able to choose between Networks 1 and 2. This increases the APUSR in RSS-based algorithm in light traffic ($\rho < 1$ in Figure 10(b)), since users



(a) Average percentage of user satisfied requests (APUSR)



(b) Blocking probability

Figure 10: MUSE-VDA performance for more demanding CBR data rate requests.

no longer join with Network 3. However, the APUSR in collective MUSE-VDA decreases, since Networks 1 and 2 cannot cover the whole area. On the other hand, all 3 networks can be used in prioritized MUSE-VDA, where the user's two sessions can be spread into multiple networks. If the bandwidth for one session cannot be satisfied, only one session will be blocked. This results in a higher APUSR and a lower blocking probability than the other two handoff schemes. Moreover, since users moving out of the limited coverage of Networks 1 and 2 cannot be served in RSS-based handoff algorithm and collective MUSE-VDA, a nonzero blocking probability can be observed all the time.

6. CONCLUSION

Expanding services through the use and coordination of diverse networks creates the challenge of developing a more complex, adaptive, and intelligent vertical handoff protocol. In this paper, MUSE-VDA has been developed to maximize the benefit of the handoff for both the user and the network. The optimizations incorporate a network elimination feature to reduce the delay and processing required in the evaluation of the cost function, and a multinetwork optimization is introduced to improve APUSR for mobile terminals with multiple active sessions. A performance analysis demonstrated significant gains in the ability to satisfy user's requests for multiple simultaneous services and a more efficient use of resources from the proposed optimizations.

In this treatment, a "proof-of-concept" has been provided based on the ability of a user or network to choose among different network types, based on different service requirements. Future work is ongoing to look at a more complex treatment, including the study of optimal and reasonable weight selection, QoS factor normalizations, and policy decision point architecture implementations.

REFERENCES

[1] I. F. Akyildiz, J. McNair, J. S. Ho, H. Uzunalioğlu, and W. Wang, "Mobility management in next-generation wireless systems," *Proceedings of the IEEE*, vol. 87, no. 8, pp. 1347–1384, 1999.

[2] K. Pahlavan, P. Krishnamurthy, A. Hatami, et al., "Handoff in hybrid mobile data networks," *IEEE Personal Communications*, vol. 7, no. 2, pp. 34–47, 2000.

[3] J. McNair and F. Zhu, "Vertical handoffs in fourth-generation multinetwork environments," *IEEE Wireless Communications*, vol. 11, no. 3, pp. 8–15, 2004.

[4] W. Wu, N. Banerjee, K. Basu, and S. K. Das, "SIP-based vertical handoff between WWANs and WLANs," *IEEE Wireless Communications*, vol. 12, no. 3, pp. 66–72, 2005.

[5] H. Badis and K. Al-Agha, "Fast and efficient vertical handoffs in wireless overlay networks," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '04)*, vol. 3, pp. 1968–1972, Barcelona, Spain, September 2004.

[6] J. Tourrilhes and C. Carter, "P-Handoff: a protocol for fine-grained peer-to-peer vertical handoff," in *Proceedings of the 13th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC '02)*, vol. 2, pp. 966–971, Lisbon, Portugal, September 2002.

[7] M. Bernaschi, F. Cacace, and G. Iannello, "Vertical handoff performance in heterogeneous networks," in *Proceedings of IEEE International Conference on Parallel Processing (ICPP '04)*, pp. 100–107, Montreal, Quebec, Canada, August 2004.

[8] H. Huang and J. Cai, "Improving TCP performance during soft vertical handoff," in *Proceedings of the 19th IEEE International Conference on Advanced Information Networking and Applications (AINA '05)*, vol. 2, pp. 329–332, Taipei, Taiwan, March 2005.

[9] S.-E. Kim and J. A. Copeland, "TCP for seamless vertical handoff in hybrid mobile data networks," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '03)*, vol. 2, pp. 661–665, San Francisco, Calif, USA, December 2003.

[10] M. Salamah, F. Tansu, and N. Khalil, "Buffering requirements for lossless vertical handoffs in wireless overlay networks," in *Proceedings of 57th IEEE Semiannual Vehicular Technology Conference (VTC '03)*, vol. 3, pp. 1984–1987, Jeju, South Korea, April 2003.

[11] Q. Zhang, C. Guo, Z. Guo, and W. Zhu, "Efficient mobility management for vertical handoff between WWAN and WLAN," *IEEE Communications Magazine*, vol. 41, no. 11, pp. 102–108, 2003.

[12] M. Ylianttila, M. Pande, J. Makela, and P. Mahonen, "Optimization scheme for mobile users performing vertical handoffs between IEEE 802.11 and GPRS/EDGE networks," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '01)*, vol. 6, pp. 3439–3443, San Antonio, Tex, USA, November 2001.

[13] M. Ylianttila, J. Makela, and P. Mahonen, "Supporting resource allocation with vertical handoffs in multiple radio network environment," in *Proceedings of the 13th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC '02)*, vol. 1, pp. 64–68, Lisbon, Portugal, September 2002.

[14] S. Sharma, I. Baek, Y. Dodia, and T. Chiueh, "OmniCon: a mobile IP-based vertical handoff system for wireless LAN and GPRS links," in *Proceedings of IEEE International Conference on Parallel Processing (ICPP '04)*, pp. 330–337, Montreal, Quebec, Canada, August 2004.

[15] M. Nam, N. Choi, Y. Seok, and Y. Choi, "Wise: energy-efficient interface selection on vertical handoff between 3G networks and WLANs," in *Proceedings of the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '04)*, vol. 1, pp. 692–698, Barcelona, Spain, September 2004.

[16] S. Jung, D.-H. Cho, and O. Song, "QoS based vertical handoff method between UMTS systems and wireless LAN networks," in *Proceedings of 60th IEEE Semiannual Vehicular Technology Conference (VTC '04)*, vol. 6, pp. 4451–4455, Los Angeles, Calif, USA, September 2004.

[17] J. M. Holtzman and A. Sampath, "Adaptive averaging methodology for handoffs in cellular systems," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 59–66, 1995.

[18] A. Mehbodniya and J. Chitizadeh, "An intelligent vertical handoff algorithm for next generation wireless networks," in *Proceedings of the 2nd IEEE/IFIP International Conference on Wireless and Optical Communications Networks (WOCN '05)*, pp. 244–249, Dubai, UAE, March 2005.

[19] A. Hasswa, N. Nasser, and H. Hassanein, "Generic vertical handoff decision function for heterogeneous wireless networks," in *Proceedings of the 2nd IEEE/IFIP International Conference on Wireless and Optical Communications Networks (WOCN '05)*, pp. 239–243, Dubai, UAE, March 2005.

[20] W.-T. Chen and Y.-Y. Shu, "Active application oriented vertical handoff in next-generation wireless networks," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '05)*, vol. 3, pp. 1383–1388, New Orleans, La, USA, March 2005.

[21] H. J. Wang, R. H. Katz, and J. Giese, "Policy-enabled handoffs across heterogeneous wireless networks," in *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA '99)*, pp. 51–60, New Orleans, La, USA, February 1999.

[22] W.-T. Chen, J.-C. Liu, and H.-K. Huang, "An adaptive scheme for vertical handoff in wireless overlay networks," in *Proceedings of the 10th International Conference on Parallel and*

Distributed Systems (ICPADS '04), vol. 10, pp. 541–548, Newport Beach, Calif, USA, July 2004.

[23] F. Zhu and J. McNair, "Optimizations for vertical handoff decision algorithms," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '04)*, vol. 2, pp. 867–872, Atlanta, Ga, USA, March 2004.

[24] R. Yavatakar, D. Pendarakis, and R. Guerin, *A Framework for Policy-Based Admission Control*, Internet Engineering Task Force, February 2006, http://www.ietf.org.

[25] D. B. Johnson and D. A. Maltz, *Dynamic Source Routing in Ad Hoc Wireless Networks*, Kluwer Academic, Norwell, Mass, USA, 1996.

[26] D. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, Englewood Cliffs, NJ, USA, 1991.

[27] K. Pahlavan and P. Krishnamurthy, *Principles of Wireless Networks*, Prentice Hall, Upper Saddle River, NJ, USA, 2002.

Fang Zhu received the B.S. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 1999, the M.S. degree from Colorado State University, Fort Collins, Colorado, in 2002, and the Ph.D. degree from University of Florida, Gainesville, Fla, in 2005. She is currently with Verizon as a Software Systems Architect. Her research interests include mobility management, quality of service, and resource management for the next generation wireless systems.

Janise McNair holds the B.S. degree and the M.S. degree in electrical engineering from the University of Texas at Austin (1991 and 1993, resp.), and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, Georgia, USA (2000). She is currently an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Florida. She is a Member of the IEEE and the ACM, and she serves on the Editorial Board of the Ad Hoc Networks Journal and the IEEE Transactions on Mobile Computing. Her current research interests are vertical handoff management, mobile user security and authentication, and medium access techniques for wireless sensor networks.